

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2002, 11:38:48 ; Search time 13.47 Seconds

(without alignments)
1859.803 Million cell updates/sec

Title: US-08-485-355B-50

Perfect score: 3374

Sequence: 1 MGDAGVNASQSPHNRGRTRNV.....GKTAARVBARRRARRARAN 647

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	130	3.9	708	1	VP40_HCMVA
2	123	3.6	886	1	VEP3_EBVA8
3	121.5	3.6	907	1	VEP3_EBV
4	119	3.5	1363	1	ILPR_BRALA
5	115.5	3.4	699	1	CH11_BACCI
6	115	3.4	1419	1	ALAL_CANAL
7	113.5	3.4	781	1	COAT_PAVNB
8	113	3.3	969	1	SACB_STRSL
9	112	3.3	1090	1	GUXB_CELFT
10	111	3.3	1902	1	P1P_LACLC
11	110.5	3.3	505	1	GUNV_ERMCA
12	109	3.2	591	1	FLGE_CAUCR
13	108.5	3.2	444	1	GUIN_ERMCA
14	108	3.2	775	1	VP4_ROT3
15	108	3.2	1260	1	ALSL_CANAL
16	107.5	3.2	2090	1	N214_HUMAN
17	106.5	3.2	709	1	HSE_SCHPO
18	106.5	3.2	776	1	VP4_ROT6
19	106.5	3.2	2201	1	POLG_CXAG
20	106	3.1	1138	1	TIFL_HUMAN
21	105	3.1	775	1	VP4_ROT1
22	105	3.1	775	1	VP4_ROT1
23	104.5	3.1	670	1	VP4_ROT1
24	104.5	3.1	1481	1	APU_THEET
25	104.5	3.1	2255	1	RRPL_SVS
26	104.5	3.1	2255	1	RRPL_SV5MR
27	103.5	3.1	1052	1	MGQC_MYGBE
28	103.5	3.1	1223	1	YMYL_CAEEL
29	103	3.1	1122	1	ADP1_MYCGA
30	102.5	3.0	776	1	VP4_ROTBU
31	102.5	3.0	2628	1	HAGA_PORCI
32	102	3.0	666	1	MOR2_ENTHR
33	102	3.0	2029	1	LAR_DROME

34	101	3.0	1216	1	NRX1_BOVIN
35	101	3.0	2514	1	POIN_SINDO
36	100.5	3.0	519	1	TRX4_HUMAN
37	100.5	3.0	562	1	NIPL_HUMAN
38	100.5	3.0	784	1	SLX15_DROME
39	100.5	3.0	1228	1	SLAP_BACST
40	100.5	3.0	2481	1	UN52_CAEEL
41	100	3.0	504	1	GUNW_ERMCA
42	100	3.0	555	1	EG10_CAEEL
43	100	3.0	1433	1	SUBF_BACSU
44	100	3.0	1625	1	CPPI_MYCTU
45	100	3.0	2832	1	NDVB_RHIME

ALIGNMENTS

```

RESULT 1
ID      VP40_HCMVA      STANDARD:      PRT:      708 AA.
AC      P16753: 069030:
DT      01-AUG-1990 (Rel. 15, Last sequence update)
DT      01-AUG-1990 (Rel. 15, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Capsid protein P40 [Contains: Assemblin (Protease) (EC 3.4.21.97);
GN      UL80 OR APNG.
OS      Human cytomegalovirus (strain AD169).
OC      Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC      Betaherpesvirinae; Cytomegalovirus.
OX      NCBI_TaxID=10360;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90269039; PubMed=2161319;
RA      Chee M.S., Bankier A.T., Beck S., Bohm R., Brown C.M., Cerry R.,
RA      Horsnell T., Hutchinson C.A., Ili, Kouzarides T., Martignetti J.A.,
RA      Predde E., Satchwell S.C., Tomlinson P., Weston K.M., Barrill B.G.;
RT      "Analysis of the protein-coding content of the sequence of human
RT      cytomegalovirus strain AD169."
RL      Curr. Top. Microbiol. Immunol. 154:125-169(1990).
RN      [2]
RP      X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF PROTEASE.
RX      MEDLINE=96399135; PubMed=8805706;
RA      Tong L., Qian C., Messariol M.-J., Bonneau P.R., Cordingley M.G.,
RA      Lagace L.;
RT      "A new serine-protease fold revealed by the crystal structure of
RT      human cytomegalovirus protease."
RL      Nature 383:272-275(1996).
RN      [3]
RP      X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF PROTEASE.
RX      MEDLINE=96399136; PubMed=8805707;
RA      Oiu X., Culp J.S., Dilella A.G., Hellmig B., Hoog S.S., Janson C.A.,
RA      Smith W.W., Abdel-Meguid S.A.;
RT      "Unique fold and active site in cytomegalovirus protease."
RL      Nature 383:275-279(1996).
RN      [4]
RP      X-RAY CRYSTALLOGRAPHY (2.27 ANGSTROMS) OF PROTEASE.
RX      MEDLINE=96399137; PubMed=8805708;
RA      Shieh H.-S., Kurumbail R.G., Stevens A.M., Stegeman R.A.,
RA      Sturman E.J., Pak J.Y., Wittwer A.J., Palmer M.O., Wiegand R.C.,
RA      Holwerda B.C., Stallings W.C.;
RT      "Three-dimensional structure of human cytomegalovirus protease."
RL      Nature 383:279-282(1996).
CC      -!- FUNCTION: THE CAPSID ASSEMBLY PROTEIN IS A COMPONENT OF THE CAPSID
CC      ASSEMBLY INVOLVED IN PROCESSING AND PACKAGING OF PROGENY DNA.
CC      AND THE CAPSID ASSEMBLY PROTEIN AT THE C-TERMINUS.
CC      -!- CATALYTIC ACTIVITY: PREFERENTIALLY CLEAVES AT ALA-|-SER OR ALA-|-
CC      ALA.
CC      -!- CATALYTIC ACTIVITY: CLEAVES -ALA + SER- AND -ALA + ALA- BONDS IN
CC      THE CAPSID PROTEIN.
CC      -!- PTM: CAPSID ASSEMBLY PROTEIN IS PHOSPHORYLATED.
CC      -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S21 (SERINE PROTEASE).

```


48 NETAADLAOSLADNTVFPANISSMPEFRMANKGKIDSDSIGWYKFKYIDPAGATES 107
 169 NITAVVR--AQGLD--VTLPLSLPTSADSNFS-----VKTEMUG--NEID-----IEC 211
 108 ARAVEYSKIPDGLVKSVDARE-----IYNECPVYDVSVPLDGRWSLSISF 160
 212 IMEDEISQVLPDGNKFNITCSGYESHPSGGLTSTSEVAT--PIPGYAVSLRLPR 269
 161 PMFR-----TAYVAANENKMSLD-----VYNDLI-----EMLNLDWRYVDS 203
 270 PYSRFLGNNISILVFSGSGPKASGDCYICQNIYFSDIPLASQDPTLTDTTYGDN- 328
 204 QMINFNDTITVYRIRVLPDYDPPTEGLVTVSDYRLTKATITCEANMPLVDQGM 263
 329 -----ATYSPMVTSE-----DANSPMVTATFM 352
 264 -----IGQYALTP----- 273
 353 AMPNNTETDCKKWTLLTSGTSCGECENISGAFASNRTFDTITVSGLTAPKTLITRTATNA 412
 274 -----SLPODYSEAYALHTLTFARPSSAALAFYWAGLPDGTAPAGPAMEQAS 324
 413 TTTTHKIVTSKAPBESTITPTLNTGFAAPNT-----TGLPSTHYPTNLTA--PAS 463
 325 SCGYLTWRHNGTTPAGSVSYLPEGFALERYDPNDG-----SWTDEASAGDVTFRQ 377
 464 TGPVTSTADVTSPTPAGTISGASP--VTPSPSRDNGTESKAPDMTSPISAVTPTTNA 520
 378 VAVDEVVYV--NNPAGGGSAPFTTVAVPPSNATNT-----VFRNLTLETRS 423
 521 TSPPTAVTTPPNATSPITGKTSPTSAVTPPNATSPPAVTPPNATITPLIGTSP 580
 424 SRRLLEMPADFGCVTANPKIEOSLKEITGCVLVHMKMNPVQLPASFGAVSN 483
 581 S-AVTPPNATSPVGEISP--QANTNHTIG-----GTSSPTVYTSPPKATSAVTG 632
 484 NPGYERTDLPYTGIR-DSPQNNSTAVAHFRSLSHSGSIVTKTYQMGVYTNVTPRG 542
 633 Q--HNTSSSTSMSLRPSISSETLSPSTSD--NSTSHMPLDLSAHPGTGENITQV--TPAS 688
 543 QRAHAGLKNELICLADLALRLGCVPAT-----DNFAAASAFANMISVLASE 595
 689 TSTH-----HVTSSPAPRGTTTQASGPNSSSTTKGGEVNTKGTGPPKN 734
 596 ATSSIISVGETAVGAOS 614
 735 ATSPQAPSGQKTAIVTVTS 753

3
 VGP3_EBV STANDARD: PRT: 907 AA.
 AC P03200: P03201:
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Envelope glycoprotein GP340 (Membrane antigen) (MA) [Contains:
 DE Glycoprotein GP220].
 GN BLUFL.
 OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 OX NCBI_TaxID=10377;
 RX MEDLINE=84270667; PubMed=6087149;
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
 RA Tuffnell P.S., Barrell B.G.:
 FT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
 RL Nature 310:207-211(1984)
 CC -!- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN

CC B-CELLS.
 CC -!- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL
 CC ENVELOPE.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC
 DR EMBL: V01555; CAA24854.1;
 DR PIR: A03763; Q0BE21.
 DR PIR: S33008; S33008.
 KM Membrane; Glycoprotein; Antigen; Late protein; Alternative splicing.
 FT CARBOHYD 47 47
 FT CARBOHYD 87 87
 FT CARBOHYD 114 114
 FT CARBOHYD 166 166
 FT CARBOHYD 169 169
 FT CARBOHYD 195 195
 FT CARBOHYD 229 229
 FT CARBOHYD 277 277
 FT CARBOHYD 318 318
 FT CARBOHYD 328 328
 FT CARBOHYD 345 345
 FT CARBOHYD 356 356
 FT CARBOHYD 378 378
 FT CARBOHYD 386 386
 FT CARBOHYD 411 411
 FT CARBOHYD 435 435
 FT CARBOHYD 443 443
 FT CARBOHYD 457 457
 FT CARBOHYD 497 497
 FT CARBOHYD 519 519
 FT CARBOHYD 533 533
 FT CARBOHYD 547 547
 FT CARBOHYD 568 568
 FT CARBOHYD 589 589
 FT CARBOHYD 610 610
 FT CARBOHYD 624 624
 FT CARBOHYD 627 627
 FT CARBOHYD 645 645
 FT CARBOHYD 656 656
 FT CARBOHYD 683 683
 FT CARBOHYD 701 701
 FT CARBOHYD 735 735
 FT CARBOHYD 746 746
 FT CARBOHYD 755 755
 FT CARBOHYD 780 780
 FT CARBOHYD 815 815
 FT CARBOHYD 858 858
 FT CARBOHYD 888 888
 FT VARSPLIC 502 698
 SQ SEQUENCE 907 AA; 94431 MW; 0750141CBAC52C9 CRC64;
 Query Match 3.6%; Score 121.5; DB 1; Length 907;
 Best Local Similarity 19.6%; Pred. No. 0.77;
 Matches 138; Conservative 66; Mismatches 266; Indels 233; Gaps 28;

48 NETAADLAOSLADNTVFPANISSMPEFRMANKGKIDSDSIGWYKFKYIDPAGATES 107
 169 NITAVVR--AQGLD--VTLPLSLPTSADSNFS-----VKTEMUG--NEID-----IEC 211
 108 ARAVEYSKIPDGLVKSVDARE-----IYNECPVYDVSVPLDGRWSLSISF 160
 212 IMEDEISQVLPDGNKFNITCSGYESHPSGGLTSTSEVAT--PIPGYAVSLRLPR 269
 161 PMFR-----TAYVAANENKMSLD-----VYNDLI-----EMLNLDWRYVDS 203

D	b	270	PVSRLGKNNLSLIFYFSGNCGKASGGGYYCQSQSNVSEDELPAQODMPTNTDTITTYGDN-	328
O	y	204	QWINEFTNDTTYVARLRVLRPLPYDVPDPDEGLVFTVS DYRLTYRAITCEANMPLTVDOGFW	263
D	b	339	-----ATYSPMYTSE-----	352
O	y	264	-----TGGOYALPT-----	273
D	b	333	AMPNNTETDFCKKPTLNSGTPSGCENISGAFASNRFFDITVSGLGTA PKTLITRTATNA	412
O	y	274	-----SLPQYDVEAAYALHTLTFEARRSSAAALAFVWAGLPQGGTAPAGTAMEQAS	324
D	b	413	TTTTTHKATFEKAPRSESTTSPILNLTGGADNTT-----TGLPSSHTPNTLTA--PAS	463
O	y	335	SGGYLTWRHNGCTTFEPAGSVSYLREGEFALERYPDNDGSW-----TDFASAGDTV	373
D	b	464	TGPTVSTADVTSPPRAGTTSGASP-----VTPSPSPMDNGTESKAPDMTSTSPVPTR	516
O	y	374	TFROYVADEVVYT-----NNPAGGSAPIFFTVARVPPSNAYINTVTR	414
D	b	517	TPNATSPRAVTTPTPNATSPPTPAVTTPTPNATSPILGKTPSAAVTTPTPNATSPILGK	576
O	y	415	N-----TLETRSSRLLELPPMPADFQGTVANNNKIKQSLKEPLGCV	458
D	b	577	TSPTSAVTTPTPNATSPILGKTSTTS-AVTTTPPNAT-GTVEETSP-QANATNRHLG--	631
O	y	459	LVSHKMRNPVQLPASSFGAVSFNNNGYERTRLDPDYTSIRDSFOONMSTAIAHFRSL	518
D	b	632	---GTSPPRPVYTSQPKNATSAVTTGQ--NHITSSSTSMKSLRRSSNPETLSPSTDNS	686
O	y	519	HSCSIVTKYTOGMEGVTVNVTPEGQFAHAGLKNELIADLADLALTRLGVVYPAT-----	573
D	b	687	HMPLLTSAHPRGENITQV--TPASISYH-----HYSTSPAPRRGTGTSQAS	731
O	y	574	---DNFPAAVSAFANMLMSVYLKSAETSSIIKSYGETFAYGAQS	614
D	b	732	GGGNSSTSTKGEVNVYTKTRPPONATSPQAPSQDKATPAVPTVS	774

```

CC RESULT 4
CC ILPR_BRALA
CC ID ILPR_BRALA STANDARD; PRT; 1363 AA.
CC AC 00246;
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Insulin-like peptide receptor precursor (EC 2.7.1.112) (ILP receptor)
CC OS Branchiostoma lanceolatum (Common lancelet) (Amphioxus).
CC OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
CC Branchiostoma.
CC OX NCBI_Taxid=7740;
CC RN 11
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=96408719; PubMed=8813726;
CC RA Pashmiforush M., Chan S.J., Steiner D.F.;
CC RT "Structure and expression of the insulin-like peptide receptor from
CC RT amphioxus."
CC RL Mol. Endocrinol. 10:857-866(1996).
CC CC -1- FUNCTION: THIS RECEPTOR BINDS TO THE INSULIN RELATED PEPTIDE AND
CC HAS A TYROSINE-PROTEIN KINASE ACTIVITY.
CC CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine -> ADP + protein
CC CC tyrosine phosphate.
CC CC -1- SUBUNIT: PROBABLE TETRAMER OF 2 ALPHA AND 2 BETA CHAINS LINKED BY
CC CC DISULFIDE BONDS. THE ALPHA CHAINS CONTRIBUTE TO THE FORMATION OF
CC CC THE LIGAND-BINDING DOMAIN, WHILE THE BETA CHAIN CARRY THE KINASE
CC CC DOMAIN (BY SIMILARITY).
CC CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC CC -1- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-
CC CC PROTEIN KINASES.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collabor
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstati
CC CC the European Bioinformatics Institute. There are no restrictions on

```

CC	use, by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL: S83394; AAB50848.1; .	
DR	HSSP: P06213; IIRK.	
DR	InterPro: IPR0006494; EGRF.L.	
DR	InterPro: IPR000719; Euk_pkinase.	
DR	InterPro: IPR003961; FN_III.	
DR	InterPro: IPR003962; FNIII_repeat.	
DR	InterPro: IPR002174; Furin-like.	
DR	InterPro: IPR002011; Receptor_tyr_kin_II.	
DR	InterPro: IPR001245; Tyr_pkinase.	
DR	Pfam: PF00041; fn3; 3.	
DR	Pfam: PF00757; Furin-like; 1.	
DR	Pfam: PF00669; pkinase; 1.	
DR	Pfam: PF01030; Recep_L_domain; 1.	
DR	PRINTS: PR00014; FNTYPEIII.	
DR	PRINTS: PR00109; TYRKINASE.	
DR	SMART: SM00060; FN3; 3.	
DR	SMART: SM00261; FU; 1.	
DR	SMART: SM00217; Tyrc; 1.	
DR	PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.	
DR	PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.	
DR	PROSITE: PS00239; RECEPTOR_TYR_KIN_II; 1.	
DR	PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.	
KW	transferase; tyrosine-protein kinase; receptor; Transmembrane;	
KW	Glycoprotein; ATP-binding; Phosphorylation; Signal.	
FT	SIGNAL	1
FT	CHAIN	30
FT		716
FT	PROPEP	717
FT	CHAIN	721
FT		1363
FT	DOMAIN	721
FT		928
FT	TRANSMEM	929
FT		949
FT	DOMAIN	950
FT		1363
FT	DOMAIN	994
FT		1283
FT	NP_BIND	1000
FT		1008
FT	BINDING	1028
FT		1028
FT	ACT_SITE	1148
FT		1148
FT	MOD_RES	1174
FT		1174
FT	CARBOHYD	51
FT		51
FT	CARBOHYD	97
FT		97
FT	CARBOHYD	137
FT		137
FT	CARBOHYD	278
FT		278
FT	CARBOHYD	483
FT		483
FT	CARBOHYD	599
FT		599
FT	CARBOHYD	617
FT		617
FT	CARBOHYD	665
FT		665
FT	CARBOHYD	666
FT		666
FT	CARBOHYD	711
FT		711
FT	CARBOHYD	732
FT		732
FT	CARBOHYD	736
FT		736
FT	CARBOHYD	743
FT		743
FT	CARBOHYD	816
FT		816
FT	CARBOHYD	885
FT		885
FT	CARBOHYD	898
FT		898
CC	SEQUENCE	1363 AA; 154104 MW; 238120B4EAB1BD65 CRC64;

```

Query Match          3.5%; Score 119; DB 1; Length 1363;
Best Local Similarity 19.8%; Pred. No. 2.1;
Matches 99; Conservative 60; Mismatches 162; Indels 180; Gaps 23.

QY 79 NMAKKIDLDSDSIGWYK-----YLDPAATESARVGEYSKIPDGLVFSYDA 128
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 416 DMDRDIDTIDEGKLFHFHPKLCRHVILTYMDKVLPEHA-----ITD----- 458
   : : : : : : : : : : : : : : : : : : : : : : : :

QY 129 EIRELYNECPVVPVSVPLDGRQKSLSFSPMRTAYVAVANVENKMSLDVYNDLIE 188
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 459 -----TDSITLITNGD-----ACGCSSTRE-----IE 480
   : : : : : : : : : : : : : : : : : : : : : : : :

```

OY	189	WLNJNLADBRYYVDSOMTINFNQDITTYVRLVRLPPTVDYDPPEGLVRYTSDVRLTJKAI	248
Dd	481	EINISKDMILL-----RMSEF-----RP-----PDRDLILTYYS-YRET-----	514
OY	249	TCEANMPRLVDGCF-WIGGOYALPTPSLPDYVSEAYALHTLTFARSSAAL-----AFW	304
Dd	515	-----EDQGLDEVDGQDAGCNTEMEKEDVSPQTANIITGLKMPQYALLVKYTK	565
OY	305	AGLRDGG-----TAPAGTAMEOASSGGYLITWRHNGTTPPAGSVSYLP	348
Dd	566	AGAREGSGAKSDIYARTDADKPRHPDQVVVYSSNSMTLLITMKRPAR--PNCNVTHYIV	623
OY	349	EGFALERDPDMSMTDEASAG-----DTYFRQVAADEVVYVTPNNGSGAPRTFYRVP	403
Dd	624	K-YKRQDEDAVEMQOREYCKGGLKPRHPQGLEDIVNNEEPNNSTIGDGTG-----CECP	678
OY	404	PS-----NAYNTPFRNMLIETTRSSRRLLEPMPRAF--GQTV----	440
Dd	679	KSEDEIRIEEAAFPQGEFENFLNNYHKKRENTRAGRRRRELPTARPFYSNQIVNVT	738
OY	441	-----ANNPKIEOSILKETLGCYLVHSMRN---PVFOLTPASSFGAVSFN	483
Dd	739	LPSTNRITVPPTFPENPNQOLETIVWNEHM--VLTGLRHFSEYILIEVIACNADAAVGS	795
OY	484	NPGYERTDLPDYTGINDSPF	504
Dd	796	GSAAVELARTQAD-----DSAD	811

ID	CHIT1_BACCI	STANDARD:	PRT:	699 AA.
CHIL_BACCI				
AC	P20533;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-FEB-1991 (Rel. 17, Last sequence update)			
DT	01-FEB-1995 (Rel. 31, Last annotation update)			
DE	Chitinase A1 precursor (EC 3.2.1.14).			
GN	CHIT1.			
OS	Bacillus circulans.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1397;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-WL-12.			
RX	MEDLINE=90366776; PubMed=2203782;			
RA	Watanabe T., Suzuki K., Oyanagi W., Ohnishi K., Tanaka H.;			
RT	"Gene cloning of chitinase A1 from Bacillus circulans WL-12 revealed			
RT	its evolutionary relationship to Serratia chitinase and to the type			
RT	III homology units of fibronectin.";			
RL	J. Biol. Chem. 265:15659-15665(1990).			
RN	[2]			
RP	MUTAGENESIS.			
RC	STRAIN-WL-12;			
RX	MEDLINE=93366760; PubMed=8103047;			
RA	Watanabe T., Kohori K., Miyashita K., Fujii T., Sakai H.,			
RA	Uchida M., Tanaka H.;			
RT	"Identification of glutamic acid 204 and aspartic acid 200 in			
RT	chitinase A1 of Bacillus circulans WL-12 as essential residues for			
RT	chitinase activity.";			
RL	J. Biol. Chem. 268:18567-18572(1993).			
CC	- I- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-			
CC	acetyl-D-glucosamine polymers of chitin.			
CC	- I SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.			
CC	- I SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL			
CC	HYDROLASES).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC	EMBL; M57601; AAA81528.1; .	
DR	PIR; A38368; A38368.	
DR	HSSP; P07254; 1CTN.	
DR	InterPro; IPR003610; Chitin_bind3.	
DR	InterPro; IPR001579; Chitinase_2.	
DR	InterPro; IPR003961; FN_III.	
DR	InterPro; IPR003962; FNIII_repeat.	
DR	InterPro; IPR001223; Glyco_hydro_18.	
DR	Pfam; PF02839; CBD_7; 1.	
DR	Pfam; PF00041; fn3; 2.	
DR	Pfam; PF00704; Glyco_hydro_18; 1.	
DR	PRINTS; PR00014; FNTYPEIII.	
DR	SMART; SM00495; CHTBD3; 1.	
DR	SMART; SM00660; FN3; 2.	
DR	PROSITE; PS01095; CHITINASE_18; 1.	
KW	Hydrolase; Glycosidase; Chitin degradation; Signal; Repeat.	
FT	SIGNAL	1
FT	CHAIN	41
FT	DOMAIN	42 699
FT	DOMAIN	42 460
FT	DOMAIN	465 549
FT	ACT_SITE	560 644
FT	ACT_SITE	204 204
FT	MUTAGEN	200 200
FT	MUTAGEN	200 200
FT	MUTAGEN	204 204
SEQ	SEQUENCE	699 AA: 73677 MW: 125982 DA: 125982

Query Match	3.48;	Score 115.5;	DB 1,	Length 699;
Best Local Similarity	21.48;	Pred. No. 1.4;		
Matches 114; Conservative	54;	Mismatches	207;	Indels 157; Gaps 25;

QY	195	DMRYV-----DSQWIMFTNDTTYVYVIRVLRPTDYVDPDTEGIVRVSQYRLTKYI	248
Db	202	DMEYVSGSDGNSKRPEKQNYTLLLSIRKRLDAAGVADCKYLLTJAASASNTYMAN	261
QY	249	TCEANMPTLVDOGFMIGGOYALTPTSLOYDVSEAV--ALHTLTFAPRESSAALAFVWA	305
Db	262	TELALIAALV-----WJ-----NIMTYDENGAMQKISAHNAPLNDPAPSA-----A	304
QY	306	GLPQGT-----APAGTPA-----ME-----QASSGGYLTWRHNGT--T	337
Db	305	GYPDANTFNVAAGAGHLDAGVPAAKVLUGVFYGRGMDGCAQAGNGQYQTCGGSSVGT	364
QY	338	FPAGS-----VSYYLPEGFALERYDPNDGSMWT-----FASAGDT----	372
Db	365	WEAGSPDFYDLBANYINKNGYT--RY-----WMDTAKVPYLYNASNKRFFISYDAEESVG	416
QY	373	-----VIFROVADEVVYVNTN-----PAGGGSAPTFYVVRP--PSNAYNTNY	412
Db	417	YKTAVYIKSKGLGAMFEMWLSGDRNKTLOKKLADLPFGTGYVPVDTJAPBSVGNARSTGV	476
QY	413	FENLT-LETRPSSRLLELPMPADFGQYVANNPKIEOSLKEFTL-----GCYLVH8K	463
Db	477	TANSTYLIANNASTDNGVY-----GYNYNCAINLATSYTGTTAITSGLTACTSTFTTK	530
QY	466	MENPVQULTPASSFGAVSEFNPNPGYERTRLDPYTGIRSEFDONMSTVAHAHFSLSHSCSI	523
Db	531	AKDAGNLSAASNAATVVS-----TQAQPGGDTQAPLAPLAPNLASTAQTSS	575
QY	524	VKRTQGMGCVYNNVTPFGQFAHAGLKKBEILLCLADLATLTGVPATYATDFFAAVVS-A	582
Db	576	ITLS-----WASTIDNCGVYGDVYNGTA-----LATTVGTGTAISGLADTSTY	621
QY	583	FAANMLSVLKSEATSI--IKSVGT--AVGAOSGLAKLPGLILSVPGK	629
Db	622	FIVKAKDAAGNVSAA5NAVSVKTAAETTNPGVSANQVNTAYTAQGLVYVNGK	673

ID	ALAI_CANAL	STANDARD:	PRT:	1419 AA.
AC	013368:			
DT	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Agglutinin-like protein ALAI precursor (Agglutinin-like adhesin).			
GN	ALAI OR A155.			
OS	Candida albicans (Yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida.			
OX	NCBI_TaxID=5476;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98053977; PubMed=9393828;			
RA	Gaur N.K., Klotz S.A.;			
RT	"Expression, cloning, and characterization of a Candida albicans			
RT	gene, ALAI, that confers adherence properties upon Saccharomyces			
RT	cerevisiae for extracellular matrix proteins.";			
RL	Infect. Immun. 65:5289-5294(1997).			
CC	-1- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.			
CC	-1- PTH: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AF025429; AAB88883.1; "			
KW	Cell adhesion; Glycoprotein; Repeat; Signal.			
FT	SIGNAL	1	17	POTENTIAL.
FT	CHAIN	18	1419	AGGLUTININ-LIKE PROTEIN ALAI.
FT	DOMAIN	399	404	POLY-THR.
FT	DOMAIN	408	418	POLY-THR.
FT	DOMAIN	437	441	POLY-THR.
FT	DOMAIN	673	676	POLY-SER.
FT	DOMAIN	687	690	POLY-SER.
FT	DOMAIN	700	703	POLY-SER.
FT	DOMAIN	719	724	POLY-SER.
FT	DOMAIN	749	752	POLY-SER.
FT	DOMAIN	787	791	POLY-SER.
FT	DOMAIN	869	872	POLY-SER.
FT	DOMAIN	875	883	POLY-SER.
FT	DOMAIN	901	911	POLY-SER.
FT	DOMAIN	1216	1221	POLY-SER.
FT	CARBOHYD	665	665	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	919	919	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1301	1301	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1326	1326	N-LINKED (GLCNAC. . .) (POTENTIAL).
QO	SEQUENCE	1419 AA:	149635 MW;	249F336F88A9D5B6 CRC64;

Query Match	3.4%:	Score 115:	DB 1:	Length 1419:
Best Local Similarity	17.7%:	Pred. No. 4.3:		
Matches 123:	Conservative	90:	Mismatches 249:	Indels 232:
			Gaps	26:
QY	39	TCGQVSPPDN--FRAAADLQASLDANVTTPPATISSPFEERNMAKKID-----	86	
Db	139	TGSSVDEDSKCFAG-----TNVTNDGSKSLIAVNEFEKSTVDOSGYLTTSRF	189	
QY	87	-----LDSISGWYFKYLDPA-----	102	
		:::	:::	:::
Db	190	MPSLINKIATLVAPQCENGYTGSTMGFEFSTGYDAIDCSNVHIGISKGVNDMNHPTSES	249	
QY	103	-GATESARAVG---EYSKIPDGLVAFSVDAETREIYNECEPVYVDVSYPLDGRQWSLIF	158	
Db	250	FSTYKSCSFEISITLYQVAVGYPGRF-IDALIS-----PSNNQYGLS--	291	
QY	159	SFPMERTIYVAVANVENKESLDVYNDLIEFLNNLADRRYVYDSEQMTNFTNDTYYVRI	218	
		:::	:::	:::

ID	COAT_PAVHB	STANDARD:	PRT:	781 AA.
DB	292	--KKNQDTCVDY-----		-MQHAFILTKMCIYKNSDAGSNI 325
QY	219	RVLRLPIYVDPDEGLVITVSDYRLTYKAITCE--ANMFTLVDOGFWIG-----		265
DB	326	VVATTRVTVDSSTA-VTLFPNPSPVDKTKTEILLIOPITTTITTSYVGVTSTYKTPAP		384
QY	266	-GOVALTPTLSLOYD-----VSEAVYALHLTTLFARPSAALAVYMGILPGCGTAPAGTPAW		320
DB	385	ICETITVIVDYVYHTTTVTYSEMTGTYITTTTRINFTDSIDTVVQVSPSPNTTTTTOFW		444
QY	321	EQASSGGLVLMRHNGTTPPAGSVSYVLDEGFALERYDPNDS-----WTFDASAGDVT		374
DB	445	SES-----FTSTTTITNSLKTGDSYIVRE-----PHNPVTVTTEFMSSEFATETIT		491
QY	375	FRQAVDEVVY-----TNNPAGGSAPTTVRVVPSNAYTN		410
DB	492	SKPEGTDSYVAREPHNPVTVTTEFMSSEYATETITNGEGTDSV---IVREPHNPVTVT		548
QY	411	TVERHTLTETRRSSRLLEPMPAPFGGTGVANNPK-IGSLILKEFLGCLVHKKRNPFV		469
DB	549	TKFWSSEYATY-----ETINKPEGTDSYVKEPNPTVTVTTEFWSSEY		592
QY	470	QLTPASSFG-----AVSFNNPGEYERTRL-----PDYIGRIDSEFQ		505
DB	593	ATTEITNGPEGTDSYVAREPHNPVTVTTEFMSSEYATETITTPGLGDSIVIHDPLEE		652
QY	506	NKSTVAVAHFRSLSHSCSYVTKYQCEGVYNNVPPFGAFHAGLKNBEILCLADDLATR		565
DB	653	SSSTY-----ALESSDSNLSAQA--ESSSSVEQ--SSSIYGLSSSDIDPLSDMPSS		702
QY	566	LTGYVPTATDNFAAASAFANMLSSVLKSEATSS		599
DB	703	STGL---TSSESTYVSIDSDSSSIESTYLSSS		733
RESULT	7			
COAT_PAVHB				
ID	COAT_PAVHB	STANDARD:	PRT:	781 AA.
AC	P07299			
DT	01-APR-1988	(Rel. 07, Created)		
DT	01-APR-1988	(Rel. 07, Last sequence update)		
DT	01-APR-1990	(Rel. 14, Last annotation update)		
DE		Probable coat protein VP1.		
OS		Human parvovirus B19.		
OC		Viruses; SSDNA viruses; Parvoviridae; Parvovirinae; Erythrovirus.		
OX	NCBI_TaxID-10798:			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-ISOLATE AU:			
EX	MEDLINE-86200451; PubMed-3701931;			
RA	Shade R.O., Blindell M.C., Cotmore S.F., Tattersall P., Astell C.R.;			
RT	"Nucleotide sequence and genome organization of human parvovirus B19			
RL	isolated from the serum of a child during aplastic crisis.";			
J. Virol.	58:921-936(1986).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on les-			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: M13178; AAC66867.1; -			
DR	PIR: A24299; VCPV19.			
DR	InterPro: IPR001403; Parvo_coat.			
DR	Pfam: PF00740; Parvo_coat; 1.			
KW	Coat protein; Glycoprotein.			
FT	CARBOHYD 46 46	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT	CARBOHYD 184 184	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT	CARBOHYD 220 220	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT	CARBOHYD 293 293	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
QO	SEQUENCE 781 AA; 86015 MW; RC6254DDB0576B07 CRC64;			

QY	112	GEYSKIPDGLYKFSVDALR----	EYNECPVYTVDSVPLDGRHMSLSIFSEPMRTAY	167
QY	112	GEYSKIPDGLYKFSVDALR---- <td>EYNECPVYTVDSVPLDGRHMSLSIFSEPMRTAY</td> <td>167</td>	EYNECPVYTVDSVPLDGRHMSLSIFSEPMRTAY	167
Db	85	GOLSDHPHALSSSSSHAEPRKENAVLSEDLHKPGQVSVLLPG-----	TNT	130
QY	168	VAVAVENKREKSLVDVNDLIEMLNNLDMRY-----	VDSQEWINTFNDT	212
Db	131	VGPCH-----ELQACRPOSADSAARHIDFRISOLAKTGINPTHTWYADELLKNIKET	186	
QY	213	TYVYRIRVLRPTVDVDPDEGLVFTVSDYRLTYKAITCEANMPTLYDQGMIGQYALTP	272	
Db	187	GFOAQV-----VKDY-FTLKG-----AAAPVAHFQ-----SL	213	
QY	273	TSLQYDVSEAYALHTLTFAPSSAALAFVWGLPQGGTAPACTPAMEQASSGGYLTWR	332	
Db	214	PEVPATYNASEKYPGMT-----SYNSAEASTGAG-----GGGSNSVKSMSE-----	254	
QY	333	HNGTTFPAGSVSYVLPBGFALERYP-----	358	
Db	255	--GATFSANSTVCTGFSRQFLRP-YDPEHHYKVFSPASSCHNMSGKRAKCTISPIMGYS	311	
QY	359	-----NDGSMTEFASAGDVTYPROVAVDEVYVTNNPAGG	393	
Db	312	TPWRYLDFENALNLFESPFLFEOHLIENYGS--IAPDALFTYISIEIAVKDY--TDKTGGG	365	
QY	394	-----SAPTFVRY--PRSNATYMTVFRNLTLET	420	
Db	366	VQYVDSITGRCLMVDYHEKRPYVLGGQDITLAPELFWYFPQYAYLVGDVNT--	421	
QY	421	RPSSRRLPLMPPADFGQYVANNPKIEOSLKETLGCY-LVHSKMRNPVQLTPASSFGA	479	
Db	422	-----QGIGDSK--KLASESAFYVLEHSS--FOLLGTGQTAS	456	
QY	480	VSFN-----NPGYERTOLDPDTGTGRDSDQNMKSTAVAHFRSLSH	519	
Db	457	MSYKPPVPYPENLBGCSOHFEYEMTNPLGSRUGVPDTLG-----GDPRFSLTH	505	
QY	520	-SCSIWTTTYOGWEGEYTVNVPTEFGQFAHAGLKNKEELIACLADLTATLTGYVA-----	572	
Db	506	EDHAIQPGNFMPGFLVNSVSTRKEDSSNTAGKALITGLSTGTSQNTPLIS-LRGPVSQPY	564	
QY	573	---TDNFAAASAPAAANMLSVLKSEATSIISKVGETAAGAA-----OSGLAKLP	620	
Db	565	HHMPTDKYVTGINAI-----SHGQTYGNAEDKEYOQGVGGRF	602	
RESULT 8				
SACB_STRSL	8	STANDARD;	969	AA.
AC	Q55242;			
DT	15-DEC-1998 (Rel. 37, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, last annotation update)			
DE	Levanucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase) (Sucrose 6-fructosyl transferase).			
GN	TFP.			
OS	Streptococcus salivarius.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
CC	Streptococcus.			
OX	NCBI_TaxID=1304;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-ATCC 25975;			
RX	MEDLINE=93322332; Pubmed=8331080;			
RA	Rathsam C., Giffard P.M., Jacques N.A.;			
RT	"The cell-bound fructosyltransferase of Streptococcus salivarius: the			
RT	carboxyl terminus specifies attachment in a Streptococcus gordonii			
RT	model system."			

```

RL   J. Bacteriol. 175:4520-4527 (1993).
CC   -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) -
CC   GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
CC   ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC   -1- SUBCELLULAR LOCATION: SECRETED. CELL-WALL BOUND.
CC   -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC   or send an email to license@isb-slb.ch).
CC   -----
CC   EMBL; L08445; AAA71925.1; -
DR   InterPro; IPR003469; Glyco_hydro_68.
DR   Pfam; PF02435; Glyco_hydro_68; 1.
KW   Transferase; Glycosyltransferase; Signal; Cell wall.
FT   SIGNAL          1
FT   CHAIN           969
FT                       LEVANSUCRASE.
SQ   SEQUENCE 969 AA: 103983 MW: D389B5B32ACF735A CRC64;

```

[illegible]

RC STRAIN-ATCC 484;
 RA MEDLINE=96003898; PubMed=7575482; Miller R.C. Jr., Warren R.A.J.;
 RA Shen H., Gilkes N.R., Kilburn D.G., Kilburn D.G., Warren R.A.J.;
 RT "Cellulohydrolase B, a second exo-cellulohydrolase from the
 RT Cellulolytic bacterium Cellulomonas flm1.";
 RL Biochem. J. 311:67-74(1995).
 RN (12)
 RP SEQUENCE OF 54-75.
 RX MEDLINE=93209933; PubMed=8458833;
 RA Melnick A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
 RT "Cellulose-binding polypeptides from Cellulomonas flm1: endoglucanase
 RT D (Cend), a family A beta-1,4-glucanase.";
 RL J. Bacteriol. 175:1910-1918(1993).
 RN (13)
 RP SEQUENCE OF 54-78.
 RX MEDLINE=94197708; PubMed=8147863;
 RA Shen H., Tomme P., Melnick A., Gilkes N.R., Kilburn D.G.,
 RA Warren R.A.J., Miller R.C. Jr.,
 RT "Stereochemical course of hydrolysis catalysed by Cellulomonas flm1
 RT Cend, a member of a new family of beta-1,4-glucanases.";
 RL Biochem. Biophys. Res. Commun. 199:1223-1228(1994).
 CC -1- FUNCTION: HYDROLYSE CELLULOSE TO A MIXTURE OF CELLOTETRAOSE,
 CC CELLOTRIOSE AND CELLOBIOSE, WITH ONLY A TRACE OF GLUCOSE. IT
 CC HYDROLYSES CELLULOSE TO CELLOTRIOSE AND CELLOBIOSE. AND
 CC CELLOTETRAOSE TO CELLOBIOSE, BUT IT DID NOT HYDROLYSE CELLOTRIOSE.
 CC HAS ALSO WEAK ENDOGLUCANASE ACTIVITY. HYDROLYSES GLUCOSIDIC BONDS
 CC WITH INVERSION OF ANOMERIC CONFIGURATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellotetraose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 CC -1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
 CC (CBD).
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY L (FAMILY 48 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L38827; AAB0822.1; -
 DR HSSP: P07986; IEXG.
 DR InterPro: IPR001919; CBD_2.
 DR InterPro: IPR003961; FN.III.
 DR InterPro: IPR003962; FN.II.Repeat.
 DR InterPro: IPR000556; Glyco_hydro_48.
 DR Pfam: PF00553; CBD_2; 1.
 DR Pfam: PF02011; fn3; 3.
 DR Pfam: PF02011; Glyco_hydro_48; 1.
 DR PRINTS: PRO0014; ENTPEP11.
 DR PRINTS: PRO0844; GLHYDRLASE48.
 DR PRODOM: PD011903; Glyco_hydro_48; 1.
 DR SMART: SM0060; FN3; 3.
 DR PROSITE: PS00561; CBD_BACTERIAL; 1.
 DR Cellulose degradation: Hydrolyase; Glycosidase; Repeat; Signal.
 KW SIGNAL.
 FT SIGNAL 1 33
 FT PROPEP 34 53
 FT CHAIN 54 1090
 FT DOMAIN 54 699 EXOGLUCANASE B.
 FT DOMAIN 700 785 CATALYTIC (BY SIMILARITY).
 FT DOMAIN 794 884 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 891 978 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 969 1090 FIBRONECTIN TYPE-III 3.
 FT ACT_SITE 513 513 CELLULOSE-BINDING (BY SIMILARITY).
 FT ACT_SITE 513 513 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 990 1089 BY SIMILARITY.
 SQ SEQUENCE 1090 AA: 114829 MW: 046B9D556F2F399 CRC64;

Query Match 3.3% Score 112; DB 1; Length 1090;

Best Local Similarity 22.2%; Pred. No. 4.8;
 Matches 130; Conservative 55; Mismatches 219; Indels 182; Gaps 29;
 QY 39 TGRQVSPDPNFTAA-----QDLASLD-ANVTFPANTISMPNPFNMAR 82
 DB 526 TGR-----PDTWNAAPPTGNDGLTVEYTSIGDDGVADTRALLFYAAKSGDTASDKAK 581
 QY 83 GKIDLDSDSIGWFKYLDPAAGATESARAVGEYSK-----IPDG----- 120
 DB 582 ALDD-----AIANNQDPLGVG-AVETRGRDYRFPDQTVANGDGIYISGWTGTPNGD 634
 QY 121 LVKFSVD-AEIRIYNEECPVYTDVSVPLDGRQMSLSTSPFER--TAYAVANVEN 175
 DB 635 VIKGVSFIDIRSFYKRD-PNMSKVQTFILDG-----AEPQFRHREMAQTAVAGA-- 684
 QY 176 KEMSLDVNDLIMLNLDMDRVVDSEQMINFTNDITYVIRLVRPYVDQDPDEGLV 235
 DB 685 -----LADYARLFDDG---TTPPTT-----APF--VPTGLQAGV 714
 QY 236 RTVSDYRLTYKATCEANNPTL-VDQGFWIGQYALPTSLPDYVSEAYALHTLTF-- 291
 DB 715 VTSEATISWTASTDTRVTVGYDVYRGATKVGATTTSTFDGCLFASVAYATVAFDA 774
 QY 292 ---ARSSAALAFVWAGLPQGTAPAGTPAMQASSGGYLTMRHNGTFPPAGSVYLP 348
 DB 775 GNVASPS--AALVTTKATPSDTPATSSSTANSVITGWSASTDNG----- 825
 QY 349 EGFALRYDPNDG---SWTFASAGDTVTFROVADEVVYTNMPAGSGAPFTYVVP 404
 DB 826 -GSGLAGYDVYRGATVAGTATLTFTDGLTASTAEYVRADVAGNVASPTANSV-- 882
 QY 405 SNAVTYVTRNTLLETPSSRRLELPPPADRGQYVANNPKIOSLKET-----LSCY 458
 DB 883 -----TTKSDTPPT--TAPSVPAALAMVTETSVALTVMNASTDYGSGLKGY 929
 QY 459 LVHSMKNRPVQLTPASSGAVSFENPG-----YERTDLDPYGIRSPQNMSTAVA 512
 DB 930 DYR-----GATRVGSTTTASTYDTGTATATAYIVRAIDNMA-----NWSAA-- 973
 QY 513 HFRSLSHSCSIYKTYO-----GWEG-----VTNVNT 539
 DB 974 -----SALSVTTKTPQTGGSCSVAYNASSWNSGFTASVRIINTGT 1014
 RESULT 10
 P1P_LACLC STANDARD; PRT; 1902 AA.
 ID P1P_LACLC
 AC P16271;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE P1-type proteinase precursor (EC 3.4.21.-) (Wall-associated serine
 DE proteinase).
 GN PRTP.
 OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
 OG Plasmid pW05.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Lactococcus.
 ON NCBI_TaxID=1359;
 OX (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=MG2;
 RX MEDLINE=88149035; PubMed=3278687;
 RA Kok J., Idenhouts K.J., Haandrikman A.J., Iedeboer A.M., Venema G.;
 RT "Nucleotide sequence of the cell wall proteinase gene of
 RT Streptococcus cremoris MG2.";
 RL Appl. Environ. Microbiol. 54:231-238(1988).
 CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
 CC GROWTH OF THE BACTERIA ON MILK.
 CC -1- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
 CC SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAS BEEN NOTED,
 CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
 CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS.


```

Db      1077  ALGNDNSAELVLTDTNMSATDQDASVKGPGSTFDL---IVNGGIPDKISS--TTTGY 1131
QY      377  GVAVEVVVTTNNPAGSGAPFTTVVPPS--NAYNTYFR-----NTLETPRSSRIELP 430
Db      1132  EA-----NTGGGTY-FTSGTYPAAYDGTNYNAGGKHNDLNTYTDAATNSTASMP 1181
QY      431  MPPADF-GGVYANNPKIEOSLLKFTLGCYLVHSKMRNFVOLTTPASSGAVSFNNPYER 489
Db      1183  YNNAAVAAOVLDYXAKHTQLKH-----EDTKYR-----LT-APFTTDLKFNNGSDOT 1229
QY      490  TRDLPDYTGI--RDSFDQNMSTAVVAHFRSLH 519
Db      1230  SEATIKVTGTYSADTKTYVNGDPTVALAQQH 1261

RESULT 11
GUNV_ERMCA STANDARD: PRT: 505 AA.
AC 047096:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
DE (Cellulase V).
GN CELV.
OS Eryinia carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_Taxid:554;
RX MEDLINE:94067016; PubMed-8246888;
RA Cooper V.J.C., Salmond G.P.C.;
RT "Molecular analysis of the major cellulase (Celv) of Eryinia
RT carotovora: evidence for an evolutionary 'mix-and-match' of enzyme
RT domains.";
RL Mol. Gen. Genet. 241:341-350(1993).
CC -I- FUNCTION: ENDOGLUCANASE WITH SOME EXOGLUCANASE ACTIVITY. THE PH
CC OPTIMUM IS ABOUT 7.0 AND THE TEMPERATURE OPTIMUM ABOUT 42 DEGREES
CC CELSIUS.
CC -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X76000; CAA53592.1; -.
DR HSSP: 085465; IASB.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF001947; CBD_3; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolyse; Glycosidase; Signal.
FT SIGNAL 1 31
FT CHAIN 32 505
FT DOMAIN 32 334 ENDOGLUCANASE V.
FT DOMAIN 335 352 CATALYTIC.
FT DOMAIN 353 505 LINKER.
FT ACT_SITE 168 168 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 505 AA; 54900 MW; DEEA337BBAD2623 CRC64;

```



```

OY 557 CIADLIATRLGCVPATDNFAAASFA-ANMLSSV-----LKSEANSSIIKSVGETAV 609
DB 489 AIFDNGVRRRLA-----QVAITAFSPNGLKGVNGNAVTVNESGYSLKASOGGA 540
OY 610 GA 611
DB 541 GA 542

RESULT 13
GUNN_ERMCA STANDARD: PRT: 444 AA.
AC Q59394;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase N precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase N)
DE (Cellulase N).
GN CELN.
OS Erwina carotovora.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATROSEPTICA FCBR C18;
RX MEDLINE=98299944; PubMed=9636315;
RA Olsen O., Thomsen K.K., Weber J., Duus J.O., Svendsen I., Wegener C.,
RA von Wettstein D.;
RT "Transplanting two unique beta-glucanase catalytic activities into
RT one multienzyme, which forms glucose.";
RL Biotechnology 14:71-76(1996).
CC -1 CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1 SUBCELLULAR LOCATION: Secreted.
CC -1 SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL: L39788; AAC37033.1; -.
CC DR HSPB: O85465; IASB.
CC DR InterPro: IPR001956; CBD_3.
CC DR InterPro: IPR001547; Glyco_hydro_F5.
CC DR Pfam: PF00942; CBD_3; 1.
CC DR Pfam: PF00150; cellulase; 1.
CC DR ProDom: PD001947; CBD_3; 1.
CC DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
CC DR Cellulose degradation; Hydrolyase; Glycosidase; Signal.
CC KW SIGNAL 1 31 POTENTIAL.
CC FT CHAIN 1 31 ENDOTGLUCANASE N.
CC FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
CC FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
CC FT SEQUENCE 444 AA; 48300 MW; FAVE4179004CBBA3 CRC64;

Query Match 3.2%; Score 108.5; DB 1; Length 444;
Best Local Similarity 20.9%; Pred. No. 2.3;
Matches 70; Conservative 48; Mismatches 138; Indels 79; Gaps 19;

OY 153 WSLSFSPMERTA--YAAVANVENKMSLDVNDLIEMLNLA-----DMRYVVDSEQW 205
DB 85 WGINFVRAVMTAAGYISNPISLANK-----VKRAVAAGSLGYIITDWHILSDNDPN 138
OY 206 INFNTDTYYIVIRLV-----RPTVDVDPPTGGLVR---TVSDYRLTKATICEANMP-- 255
DB 139 IYKAAKTFEFEMAGLYSSSPVIVIEIANEPNGVTWNGQIRPVALEV-TDITIRSKDPDN 197

```

```

OY 256 -TLVDGFMIGGQVALPTPSLPQVDVSEAYALH-----TLFARPSAALAF 302
DB 198 LIIVGTGTWSDIHDAADNQLP--DPNTLYALHAFYAGTHGCFRLDRIDYRSGAIIIVS 255
OY 303 VWAGLPGGGTAPAGTPAMEOASSGGYL-----TW-----RHNGTTPAGSVSYLPEGFALE 354
DB 256 EW-----GTSDA-----SGNGGPFLEPSQWTWIDFLNNRGVSWVNWSLTLKSEASALA 303
OY 355 KYDPNDGSWTD--FASAGDTVTFFROYAVDEVVYVNNPAGGSAPFTYVRV-----PSNA 407
DB 304 PGASKSGWTEQNLSTSGKFVR-----EQIRAGANIGGGDPTP-TPTTPTPTPNPGNG 355
OY 408 YNTVY---FRNTLTPRSSRLLELMPPADGQT 439
DB 356 TTGDVVLQYRN--VDNPSDDAIRMAVIRKNGST 388

RESULT 14
VP4_ROT3 STANDARD: PRT: 775 AA.
AC P39033;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer capsid protein VP4 (Hemagglutinin) [Outer layer protein VP4]
DE [Contains: Outer capsid proteins VP5 and VP8].
GN S4.
OS Human rotavirus (serotype G3 / strain AU-1).
OC Viruses: dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=39013;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92356070; PubMed=1322955;
RA Isegawa Y., Nakagomi O., Nakagomi T., Ueda S.;
RT "A VP4 sequence highly conserved in human rotavirus strain AU-1 and
RT feline rotavirus strain FRV-1."
RL J. Gen. Virol. 73:1939-1946(1992).
CC -1 SUBCELLULAR LOCATION: Outer capsid.
CC -1 PWM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
CC OTHER PRODUCT IS VP5.
CC -1 SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL: D10970; BA001747.1; -.
CC DR PIR: J01638; J01638.
CC DR InterPro: IPR000416; Cap_VP4.
CC DR Pfam: PF00426; VP4; 1.
CC KW Coat protein; Glycoprotein.
CC FT CHAIN 1 775
CC FT CHAIN 1 241
CC FT CARBOHYD 17 775 OUTER CAPSID PROTEIN VP4.
CC FT CARBOHYD 17 17 OUTER CAPSID PROTEIN VP5.
CC FT CARBOHYD 32 32 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 183 183 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 237 237 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT SEQUENCE 775 AA; 87092 MW; 41480DC91A89B18A CRC64;

Query Match 3.2%; Score 108; DB 1; Length 775;
Best Local Similarity 20.9%; Pred. No. 5.6;

```


! : | |||
Db 968 TSNQDNQSG 976

Search completed: May 29, 2002, 11:40:45
Job time: 117 sec
